The Effects of SSD Caching on the I/O Performance of Unified Storage Systems

Heidi Sandoval
CSU Bakersfield

Matthew Dwyer
Lynchburg College est. 1903

Anthony Pearson
St. Cloud State University

Abstract

Increases in computing power that have greatly enhanced high performance computing have forced large-scale supercomputers to upgrade to faster storage systems. We explored the combination of a Unified backing Storage System and SSD/Flash caching as a means of handling the large amounts of data that will be archived in future HPC systems. Using the dm-cache and bcache algorithms, caching was implemented, and I/O performance on the storage system was benchmarked. A connector server communicated with the storage servers and produced a block device that uses their available cloud storage space. A SCSI block device was used by the initiator server, and a POSIX file system was created on this remote block device that allowed for benchmarking. I/O bandwidth, among multiple devices and caching methods, was tested on our test cluster. Solid state drives and Flash were used as burst buffers to speed up the Unified Storage System's read and write performance.

Overview

Goal:
Implement a Unified Storage System and test the I/O performance with and without SSD/Flash caching.

Steps:

1. **Setup an object based storage system**
   a. **Erasure coded to improve reliability**
2. **Layer a POSIX style interface over this system**
   a. **Using SCSI Fibre Channel Protocol (FCP)**
3. **Use SSD/Flash caching to increase performance**
   a. **bcache**
   b. **dm-cache**
4. **Use benchmarking tools to test performance**
   a. **IOZONE**
   b. **dd**

Testbed

Figure 2, shown below, provides a graphical representation of the testbed. The testbed includes nine servers which include:

- **Supervisor/Login**
- **Connector/SCSI Target**
- **SCSI Initiator**
- **6 Storage Servers**

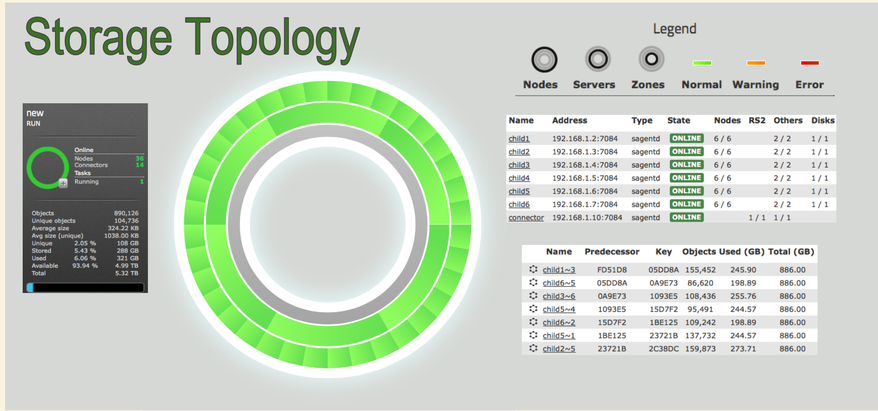

Figure 1 (Above) depicts the topology of the cloud storage system. This system is comprised of 6 storage servers, 1 connector, and a supervisor. The storage servers contain one logical drive made up of two physical HDDs. Each logical drive is configured to have 6 nodes that house data objects. The nodes are randomly distributed within the storage system to maximize its reliability.

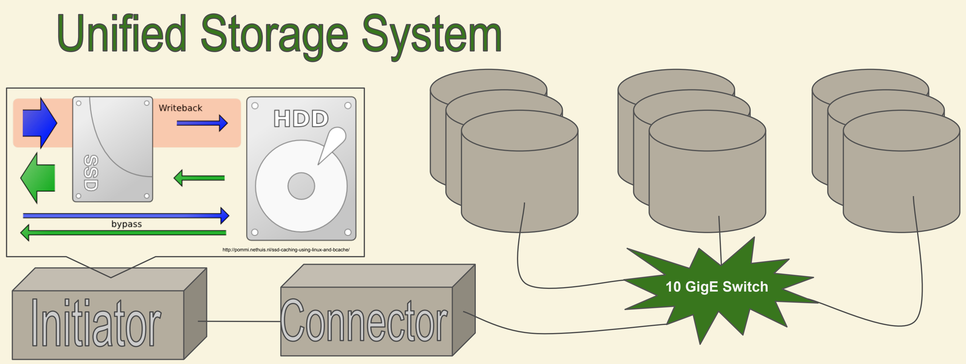

Figure 2: The 10 Gb Ethernet switch connects the Storage Servers, Supervisor, and Connector together. It serves as the cloud storage system's primary method of communication, and it allows the Supervisor and Connector servers to access the cloud. Using a FUSE mount, the Connector creates SCSI hard drives that are visible by the Initiator through Fibre Channel. Those SCSI drives are used as origin/backing devices for caching algorithms as a means of improving the I/O performance of the storage system.

Conclusions

We successfully set up a Unified Storage System that used two different caching schemes: bcache and dm-cache. We found that:

- **File sizes must be large enough to avoid internal caching (via RAM)**
- **Current versions of software provided are underdeveloped and as such unstable**
- **Kernel panics, which interrupted testing, may have occurred due to FUSE misconfiguration**
- **Our SCSI target devices are often undiscoverable**
- **The SCSI fibre channel protocol operates with a limited number of Linux distributions**

With the extent of these obstacles, we find that it is too early to accurately enact the cloud object storage system using the provided software.

Future Work

- **Analyze changes in I/O performance when caching in different areas of the Unified Storage System**
- **Evaluate additional caching methods and fine-tune existing techniques**
- **Group multiple Flash devices in a RAID 0 array**
- **Expand the cloud storage system to include a separate metadata ring**
- **Unified storage setup needs further investigation**

Acknowledgments

Mentors: H.B. Chen, Sean Blanchard, Jeff Inman
Instructor: Dane Gardner assisted by Chris Moore

ISTI Information Science & Technology Institute
Los Alamos National Laboratory EST.1943
NSF
PRObE Parallel Reconfigurable Observational Environment
New Mexico Consortium
Chartreuse

Contact Information
Anthony Pearson - pean0906@stcloudstate.edu
Matthew Dwyer - dwyer_m@lynchburg.students.edu
Hdidi Sandoval - heidi.sandoval@hotmail.com

LA-UR-14-26015